RMLE Online
Research in Middle Level Education

2013 • Volume 37 • Number 4 ISSN 1940-4476

Promoting Rigor-in-Practice through School Self-Evaluation: A Middle School's Experience with Model Development, Implementation, and Evaluation

Michael J. Mann
School of Public Health
West Virginia University
Morgantown, WV

Megan L. Smith
College of Education and Human Services
West Virginia University
Morgantown, WV

Introduction

For years, the No Child Left Behind Act (NCLB) has demanded that our nation pay ever closer attention to school evaluation (Schmidt, 2008; Stanik, 2007). Never before have the stakes of evaluation been so high. School and district jobs, funding, and local control of educational policy and practice are all more directly jeopardized by unfavorable evaluations than at any other time in our country's history (Schmidt, 2008; Stanik, 2007). These high-stakes consequences demand that educators be evaluation experts, fully prepared to take part in the debate about U.S. education and school evaluation.

Choosing which objectives are evaluated and which are not is a particularly contentious point in this debate, especially when considering the unintended and often negative impact these decisions can have on schools. High-stakes NCLB school evaluation has successfully focused national attention on what's being evaluated—math and reading scores, initially—but may have done so only at the expense of other top priorities (Stanik, 2007). Examples of these important, but now frequently overlooked priorities, include efforts to meet the broader developmental needs of the whole child, and school social services designed to support struggling or disadvantaged students.

Although many examples show that heightened attention to NCLB evaluation has directly contributed to some schools' redoubling of their efforts and successfully improving student reading and math scores, many equally valid examples can be provided in which focus on NCLB outcomes has resulted in a diminishment of other critical parts of the curriculum or school services (Stanik, 2007). Many of these priorities may have been rightly determined to be outside the scope of national evaluation efforts but are still deserving of adequate attention and resources from local stakeholders.

This article describes one middle school's efforts to use model building and school self-evaluation to

hold itself accountable for its own priorities. These priorities were established by local administrators, teachers, and students, in addition to those established by NCLB. Further, this article discusses this school's efforts to make evaluation a more valuable and meaningful part of the work of all school professionals—particularly their efforts to use evaluation to enhance outcomes related to the whole student—and to develop a strong "rigor-in-practice" culture within the school. It concludes with a discussion about the role of evaluation in middle schools and recommendations for evaluators interested in implementing similar practices.

School profile and background information

The Sunshine Learning Academy (a fictitious name) is an alternative middle school located in a suburban community in the southeast United States. The school serves approximately 200 students and families. Most students are referred to the Sunshine Learning Academy (SLA) after multiple suspensions, an expulsion related to behavior problems in school, or legal violations in their community. Some of these students are chronically truant and need to complete middle school at an accelerated rate before resuming their educational career in a high school placement. More than 70% of SLA students receive free or reduced-price lunches, and approximately 60% have a special education designation. Frequently, SLA students' behavior problems are accompanied by a history of academic failure. SLA's goals focus educators' efforts on helping students adopt the attitudes, behaviors, and academic skills necessary for successful reentry into traditional public middle schools.

If they are to achieve their goals, SLA educators' work cannot be limited to improving math and reading outcomes. For most SLA students, accomplishing a set of social or developmental objectives often precedes academic achievement. Some of these objectives include learning effective impulse control, the ability to delay gratification, pro-social values, communication skills including de-escalation and negotiation strategies, goal-setting skills, the ability to cope with a host of intense stresses and traumas, and the recovery of an often hobbled sense of self-confidence.

Although SLA has enjoyed a 12-year history of producing generally positive student outcomes, faculty and administrators became curious about the possible causes or correlates of various peaks and valleys in the school's performance over time. As a result, this author was hired to help the school accomplish four main goals. These goals included:

1. Develop a plain-language, collaborative working model that accurately described the elements SLA personnel and students believed to contribute most to student success in the alternative school environment.
2. Ground the model in theoretical and evidence-based research.
3. Examine the school's systems and incentive structures in order to realign them to better support model implementation.
4. Develop an evaluation plan designed to improve student and school outcomes.

Phase 1: Model development

Psychological theory suggests that people behave and make decisions using mental or working models (Resick, et al., 2010; see also Kennedy & McComb, 2010). These models consist of conscious and unconscious explanations of how the world works. They are highly personal and reflect a person's best efforts to integrate her or his life experiences into a narrative the individual can use to make decisions. Generally speaking, whether or not they are objectively rational, these models reflect basic assumptions about cause and effect and, as a result, exert a tremendous influence over the behaviors a person chooses.

Making these models explicit creates a tool or marker that can help organize diverse members of any group. A shared model clarifies the intended outcomes of the organization and provides group members with a common explanation of how the organization works to achieve those outcomes. A good model helps the group understand what is important and which behaviors most contribute to their collective definition of success (Kennedy & McComb, 2010; Resick, et al., 2010).

The most powerful models resonate deeply with as many members of the group as possible. Many pathways lead to this end, including hiring members whose personal models are similar to the organization's model, training and developing inclusive programs that make employees feel valued as a group, and—especially for existing groups—developing a collaborative model that accurately reflects and reinforces the often unspoken assumptions of the group.

SLA educators chose to develop a collaborative model based on what they believed most contributed to the school's periods of greatest success. This model was developed in several phases. In the first phase, interviews were conducted with each teacher, staff member, and administrator while a series of focus groups were conducted with both past and current students. Private, one-on-one interviews were conducted with each school teacher, staff member, and administrator, with the assurance of confidentiality. Student focus groups were conducted off campus at a nearby public park. These focus groups intentionally included both students who were currently completing the alternative school program successfully and those who were struggling with successful program completion. Additionally, a focus group comprised of recent program graduates (less than one year post-graduation) included both students who were successfully completing mainstream public high school and those who were struggling in their high school placements. In all interviews and focus groups, participating students and adults were encouraged to speak candidly as experts who had meaningful opinions about and experiences with what both contributed to and constrained the school's effectiveness.

During the second phase, based on the feedback provided by school personnel and student groups, the first draft model was developed by the author and presented to the school's administrative team. During this presentation, the administrative team suggested minor revisions that were incorporated into the next draft of the model. After administrative review of these revisions, the revised model was presented to school personnel, and further revisions and refinements were suggested by various faculty and staff members. This process of drafts and revisions was repeated two additional times, until school personnel at all levels agreed on a draft of the model ready for student comment.

In the final phase, the revised model was presented to student focus groups. Again, these focus groups included both current and former students experiencing a range of positive and negative program outcomes. These focus groups emphasized the short-term outcome and long-term impact portions of the model. Because these portions of the model emphasized desired student outcomes, it was particularly important to determine whether or not these portions of the model were easily understood by students and made sense to them as outcomes important to their current and future school success. The initial model development plan included taking student comments and suggestions for revisions back to school personnel for a final phase of model revision. However, students described the model as a tool that simplified and clarified the expectations of school personnel; students further conceptualized the model as a concise representation that focused on helping them prepare for success in their high school placement. In fact, student feedback was favorable enough that the student focus groups did not result in substantive recommendations for revision.

In all phases with all groups contributing to this successful model, materials were sent to participants (and thus to each related group) approximately one week prior to any feedback meeting; as a result, all meetings were kept to no more than 35 minutes in length. Throughout this process, the author worked to integrate the experiences and opinions of students and school personnel with evidence-based learning and behavior modification theories as well as the most current brain science and developmental research. The final model included concepts and relationships that were firmly supported in the findings of the research literature.

Together, these efforts resulted in the development of the *Model of Empowered Student Behavior for Alternative Middle Schools*, illustrated in Figure 1. A brief description of the model is provided as an introductory example of modeling.

Example 1: Describing SLA's process-context related school objectives. The process-context portion of a model describes the processes and contexts that best support achieving the model's desired outcomes and impacts (Chen, 2005; McKenzie, Neigler, & Smeltzer, 2005). This portion of the model emphasized how SLA educators planned to influence behavior change and academic improvement in at-risk middle school students and, as such, described the processes and contexts believed most likely to create an environment in which these changes would occur. A brief description of this portion of their model is presented as an example.

The *Model of Student Empowerment for Alternative Middle Schools* begins by emphasizing the importance of a predictable and well-structured school environment rich in adult role models. Predictable school structures both minimize conflict and, when conflict does occur, clarify the cause and effect relationships between student behavior choices and rewards or consequences (Ames, 1992; McClure,

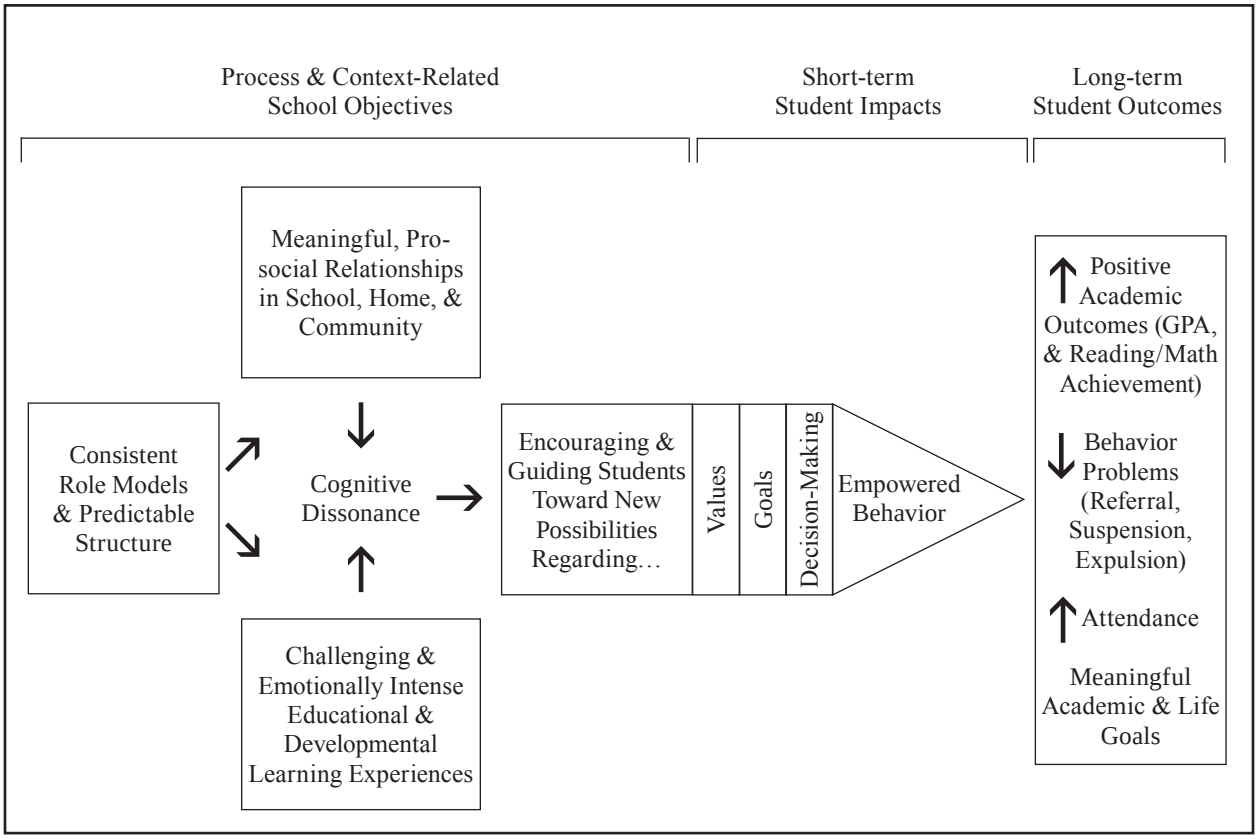

Figure 1. Model of Empowered Student Behavior for Alternative Middle Schools

Yonezawa, & Jones, 2010; Taylor, 2009). Role models are living examples of the benefits of reasonable pro-social behavior. Role models are important because students are more likely to adopt behaviors they have seen succeed (Bandura, 1977). According to this model, being a role model means the adults in the school successfully exhibit the behaviors they are asking students to choose.

Next, the model suggests that behavior change takes place in the context of meaningful relationships and challenging, emotionally intense learning experiences. Often, young people find the strength to grow or change in the context of relationships (Glasser, 1999). Supportive relationships engage young adolescents; when a respected or influential adult believes young adolescents can grow, young people often make efforts based on or bolstered by that belief. Additionally, at-risk adolescents often rate their potential, especially their academic potential, far lower than it actually is (Multon, Brown, & Lent, 1991). Authentically challenging and emotionally intense learning experiences, especially those with successful outcomes, encourage students to reconsider their capabilities and expand their possibilities for the future. Further, because the limbic portion of the brain is often more developed than the cognitive-control portion in early adolescence, the emotional intensity of relationships and experiences can make a substantial contribution to the subjective influence of both factors on student growth, development, and learning (Dahl, 2001; Keating, 2004; Sowell, et al., 2002).

Finally, the ultimate goal of this portion of the model centers on cognitive dissonance, which is created when our expectations for ourselves are challenged and we are forced to reconcile that challenge (Forgas, 2001). For instance, when a young person who believes she/he cannot be successful in math actually experiences true success, the student is forced to reappraise her/his assumptions about her/his self, and, with proper guidance, this young adolescent has the opportunity to choose increasingly pro-math attitudes and behaviors. The SLA program intentionally creates the conditions in which this type of success, and its accompanying dissonance, is likely and encourages students to expect increasingly higher levels of achievement from themselves.

As this brief example illustrates, the process-context portion of a model describes both the context in which

the best outcomes are most likely and the processes that most contribute to those outcomes. Identifying and describing a school's essential process-context objectives helps establish a clear road map to success for all school personnel. In broad terms, these objectives clarify which types of activities are to be conducted at the school and how they are to be conducted, if the school is to achieve its goals.

Example 2: Describing SLA's short-term student impacts. The short-term student impact portion of a model identifies the intended immediate impacts of the process-context portion of the model. Often, these impacts consist of mediating variables that influence the achievement of longer-term goals. In SLA's case, this portion of the model describes how changes in student intrapersonal variables contribute to the school's longer-term influence on students. Specifically, it suggests school personnel help students resolve cognitive dissonance in a way that bolsters four key mediating variables: values, goal setting, decision making, and empowerment. According to the model, young people who demonstrate pro-social values tend to set more positive and productive life goals (Barnstable, Cargill, Gehlbach, & Workman, 1997). If these goals are personally meaningful and relevant to the student, student decision making will be strongly influenced by them and will lead to increasingly pro-social, success-oriented behavior. In turn, these successes reinforce student commitment to their newly established values, goals, and decision-making skills while promoting increasingly empowered behavior (Barnstable, et al., 1997). This process creates a positive feedback loop that promotes continually deepening motivation and commitment to the long-term goals of the model.

Achieving these short-term student impacts is critical because they precede and mediate overall student success. As such, improvements in these four mediating variables provide a foundation for further necessary changes in student behavior. The scientific literature addressing at-risk student achievement strongly suggests that students scoring higher in each of these areas—especially empowerment or self-efficacy—are more likely to reduce their rates of problem behavior and improve their academic performance, school attendance, graduation rates, and goal-setting behaviors (Moriarty, Douglas, Punch, & Hattie, 1995; Multon, et al., 1991; Pajares & Graham, 1999).

This brief example demonstrates the importance of accurately describing short-term impacts. These objectives help school personnel direct all process-context activities toward predetermined ends—impacts likely to mediate or provide a foundation for the longer-term student outcomes identified as the school's ultimate goals.

Example 3: Describing long-term student outcomes. The long-term student outcomes portion of the model identifies the end goals or products of the school or program. Ultimately, SLA's primary goals focus on helping at-risk middle school students develop the attitudes, behaviors, and academic skills necessary to successfully reintegrate into public school. While the previous two portions of the model describe what the school does to promote growth or change, the long-term student outcomes portion of the model describes the desired changes themselves. For example, goals included reduced behavior problems, improved school attendance, improved GPA and reading and math scores, and clear school and life goals. All other model objectives and components are designed and expected to contribute to achieving these ends.

Other model considerations. Choosing model components requires a careful balancing act. This balancing act demands that no essential concepts are excluded while no unessential concepts are included. The best working models are as parsimonious as possible (Nowak, 2004). Well-constructed models only include the elements absolutely necessary for predicting a successful outcome. This parsimony makes the model manageable and practical, in terms of both daily use and evaluation measurement. Obviously, any model could be expanded to include every possible factor. In practice, this results in unwieldy models that are hard to use in routine decision making and are equally difficult to measure.

For SLA, the *Model of Empowered Student Behavior for Alternative Middle Schools* provided a clear representation of what school personnel and students believed most contributed to school success. This level of clarity made it possible for SLA personnel to be confident about which priorities were most important at the school and how to best pursue those priorities. The model was well supported by current research and presented scientifically rigorous concepts in plain, accessible, and user-friendly language. As a result, the model came to serve as the primary reference point for all school planning; training; decision making; and, ultimately, school self-evaluation and initiatives designed to promote increasingly rigorous practice.

Phase 2: Model and systems alignment

Systems promote the efficient and consistent implementation of repetitive tasks within any organization. For example, every school uses some type of system for reporting attendance, determining and dealing with behavior problems, and recording and tracking student academic progress. These systems organize the routine work of the school and institutionalize methods for reliably achieving desired outcomes.

Systems also exert values and priorities within the school environment (Abreu, Macedo, & Camarinha-Matos, 2009). For instance, NCLB represents a system of evaluation and accountability that some believe values multiple-choice standardized test scores more than other methods of assessment, such as course grades, student portfolios, or project-based assessment. Whether or not this assessment of NCLB is correct or fair, it is difficult to argue that NCLB does not exert a set of priorities that influence how schools operate.

Once SLA developed a working model accurately depicting its desired processes and outcomes, all relevant systems had to be examined and adjusted to support the reliable implementation of those processes and the consistent achievement of those outcomes. Models are only useful when they are widely used. If approached purposefully, systems can institutionalize a school's values in healthy and positive ways, and because they are so ingrained in the daily work experience, few things support consistent use better than well-conceived and implemented systems.

Examining systems can be exceptionally difficult (Miller-Williams & Kritsonis, 2009). Old systems are so entrenched in how school personnel operate that it can be difficult to imagine working in any other way. This tendency represents both a strength and a weakness of the systems approach. When systems are congruent with the organization's highest values, they make choosing effective processes almost effortless. When they are not congruent, they can subtly undermine a school's effectiveness without prompting anyone to scrutinize their influence.

For example, helping students learn to choose pro-social behavior represented one of the most central and important goals at SLA. Over the past decade, the school developed elaborate behavior modification systems central to its daily operations. These systems had become so ingrained in the culture of the school that they had not been reevaluated in many years.

However, after SLA developed their working model, they took the time to reexamine each of these systems and determine whether or not they needed to be adjusted and better aligned with their newly explicit values, processes, and outcomes. SLA's original behavior modification program consisted of four basic systems. These included systems for providing each of the following: (a) student feedback, (b) behavior counseling, (c) student incentives and disincentives, and (d) exit criteria for their departure from alternative school placement and reentrance into the traditional public school environment. Examples of adjustments made to two of these systems are described below.

Example 1: Aligning the student feedback system. Prior to model development, SLA used a point-based feedback system. In this system, each student carried a point card. Each class period students were scored on six behavior categories and awarded points based on how well they met the school's expectations for each category. Totaled, these points indicated whether or not a student's behavior was "on-target" for the class period; and when all periods were totaled, for the day and, eventually, the week. Program exit was tied to the accumulation of points over time.

When SLA personnel reexamined this system, they discovered several ways in which it was not congruent with their model. For instance, each teacher invariably scored points differently, which made the school structure less predictable. Also, teachers spent up to a quarter of a given academic class period scoring point cards, which detracted from the time students were engaged in challenging learning experiences. Further, because points were so closely tied to student exit, students receiving fewer points frequently responded in intensely negative ways. Often, the ensuing power struggle strained the teacher-student relationship. Finally, because student responses to losing points—and by extension having their exit from the alternative school placement postponed—could be so emotionally charged, teachers were constantly tempted to score underperforming students more highly than they should, preventing experiences with cognitive dissonance at the heart of the model.

After adjustments to this system, students no longer carried point cards. The behavior expectations once scored on the point cards were posted in each classroom. On-target behavior was reinforced verbally throughout the class period. Only students performing above or below these expectations were formally addressed during a class period.

Students performing above expectations received a commendation. Students who failed to meet these expectations received a warning and, if they continued to perform below expectations, received a behavior referral, which had the potential to delay student exit. Students perceived this system as more forgiving and appreciated the opportunity to discuss their situation before a referral had an impact on their exit date (Mann, 2010) .

The impact of the newly adjusted system was immediate. Students described feeling less defensive, having better relationships with teachers, and being more open to teacher feedback. Verbal praise for on-target behavior was perceived as being more meaningful than a point card score. Further, detaching teacher feedback from points helped students feel this feedback was more authentic, caring, and less punitive. Unknowingly, the process of providing every student with written behavior scores every class period had contributed to creating an environment of constant scrutiny. After this adjustment, the atmosphere within the school relaxed substantially. Both teachers and students described less student behavior incidents, more positive teacher-student relationships, more time invested in learning, improved inter-teacher consistency, and a more predictable school structure—all outcomes better aligned with the new model than the outcomes produced by the previous system (Mann, 2010).

Example 2: Aligning the student exit system. SLA's student exit system also required adjustments. Previously, SLA used a level system to determine when students were ready to transition back into traditional public schools. Movement between levels was determined by the accumulation of points from point cards and accomplishing a list of other requirements, ranging from completing community service hours to having one of the "Top 10" highest point card totals for a given week. Actual promotion to higher levels required a vote from school personnel. Higher levels required increasingly unanimous votes.

When compared to the model, three fundamental problems arose with this system. First, keeping track of the system was complicated. Errors were made often, and even when they weren't, the system was so complex that students rarely knew how they were progressing in the exit process. This negatively influenced the predictable structure of the school and caused tension that strained teacher-student relationships. Second, none of the exit criteria were directly related to any of SLA's desired long-term student outcomes. Instead, indirect symbolic and representational goals were used—levels and points. This made it needlessly difficult for students to see clear connections between the behaviors required for exit and their progress toward achieving that exit. Third, because of the symbolic nature of the exit criteria, teachers were required to constantly make decisions and assign values that impacted student exit. This approach made it too easy for students to make the argument, sometimes convincingly, that they were failing to exit because teachers did not assign the appropriate points. Using indirect and inauthentic mediators of points and levels allowed students to be less directly responsible for their exit than did using more direct or authentic mediators.

After reviewing this system, SLA decided it required extensive adjustments to adequately fit the new model. After several revisions, the new student exit system was completely reoriented to track student progress toward achieving the model's long-term student impacts while attending SLA. Upon school entry, each student's earliest date of exit was established. Between entry and that date, student exit was determined by achieving desired rates of school attendance, behavior referrals, grade point average, and setting individual academic and life goals. To exit on time, SLA students had to maintain a 95% attendance rate; average no more than two minor behavior referrals per month, with no major referrals during the month preceding exit; earn a grade point average of 2.5 or higher; and be able to list three academic or life goals and the steps required to achieve those goals. Student progress toward these outcomes was calculated and provided to each student weekly.

The two main advantages associated with these adjustments included the new system's comparative simplicity, and perhaps more important, its being based on what students perceived as more authentic outcomes (i.e., behavior referrals, attendance rates, grade point average, academic/life goals and plans). Students easily understood these behaviors would be required of them back in their traditional school placement, and they more easily saw themselves as directly responsible for achieving or failing to achieve these desired outcomes. They were used to their teachers' role in tracking attendance, behavior, and grade point average and knew teachers would be doing this in their post-alternative school placement. No symbolic and indirect system clouded whether they were or were not responsible for their own outcomes. As a result, students who consistently

demonstrated these improved outcomes while attending SLA were more likely to demonstrate the same behaviors in their next public school placement.

Other system alignment considerations. Not all systems required adjustment. Many systems were already successfully aligned with the model. For instance, SLA's student incentive and behavior counseling systems were already congruent, and no adjustments were necessary. Modeling and systems realignment actually strengthened commitment to these preexisting systems, as school personnel were more able to describe how and why these systems already contributed to student and school success.

Also, adjustments to the two systems described above simplified the school's work. Committing to rigor-in-practice through self-evaluation does not have to mean increasingly complicated, more demanding, or more financially-strained work. Implemented effectively, these activities can help schools determine what is not essential, eliminate less effective strategies, and streamline workloads.

Finally, in the best situations, schools will reexamine all of their systems. Including how they coordinate annual planning and the school calendar; curriculum, unit, and lesson planning; job descriptions and performance reviews; and training and staff development. The better job a school or district does in aligning each of these systems with its model, the better job its educators will do of operating as congruently as possible with all of their goals and achieving the best possible results.

Phase 3: School self-evaluation

True evaluation focuses on improvement (Chen, 2005; McKenzie et al., 2005). The best evaluations provide feedback about a school or student's progress toward previously established goals or objectives; identify areas that require further attention, growth, or improvement; and identify areas of excellence while reinforcing commitment to efforts and processes producing positive results. When used properly, evaluation helps committed professionals systematically reflect on their performance and strategically plan for increasing success (Chen, 2005; McKenzie et al., 2005). Evaluation represents a critical skill to be used by professionals dedicated to increasingly rigorous practice.

In the best case, school evaluators strive to be objective advocates for the school, far enough removed from day-to-day operations to provide feedback about how well the school is achieving its desired outcomes while remaining detached from personal motives or political agendas. Also in the best case, school personnel initiate the evaluation process themselves as they pursue their vision of excellence and success (Chen, 2005; McKenzie et al., 2005). Simply put, evaluation need not be intrinsically punitive or designed to catch people underperforming. Evaluations can be positive and supportive and can kindly and respectfully encourage the success of every member of the school community.

Many educators, however, respond reluctantly—if not fearfully—to evaluation efforts (Schoen & Fusarelli, 2008). At least four fundamental problems contribute to this response. First, educators may wrongly equate high-stakes, one-size-fits-all, accountability-oriented programs like NCLB with evaluation in general. Second, educators resist having their personal or institutional efforts reduced to a set of over-simplified numbers or scores that do little to account for context or circumstance. Third, evaluations may be conducted by evaluators who lack the experience, professional credibility, modeling, reciprocity, or interpersonal skills required to help school personnel be confident their situations and efforts will be adequately understood and accurately represented. And fourth, educators may have had negative experiences with evaluation that undermine their resolve to participate again. In particular, experiences in which evaluation felt like busy work or something empty and unrelated to candid dialogue that made little or no contribution to the success of the school.

At SLA, educators initiated their own evaluation program. Although they contracted an outside evaluator to facilitate the process, ownership concerning these efforts remained at the local school level. They made this choice based on an organization-wide commitment to school improvement that included NCLB objectives but also supplemented them in ways that were important to members of the school. Examples of SLA's evaluation and reporting methods are presented below.

Example 1: Evaluating SLA's process-context objectives. Process-context evaluation answers the question "Did the school consistently and adequately implement the processes and create the contexts the model suggests contribute to successful outcomes?" (Chen, 2005; McKenzie et al., 2005). In SLA's case, these processes included providing role models and predictable school structure, promoting meaningful and pro-social relationships at home and

at school, facilitating challenging and emotionally intense learning experiences, and using the ensuing experiences with success to help students consider increasingly pro-social academic and behavioral possibilities. Each of these core processes was evaluated using three primary methods: student interviews; staff interviews; and observations, including records review.

Student and staff interviews were conducted using an interview script developed specifically for the SLA evaluation. Interview questions corresponded directly to the processes-context objectives found in the model and the evaluation criteria established for each (see Table 1 for example criteria). Notes were taken describing interviewee responses. All responses were combined and analyzed for trends and themes. Each year, 30 to 40 student interviews were conducted, and each teacher and school administrator was interviewed twice. Interviews intentionally were designed to be brief. The average interview took 15 to 20 minutes to complete.

Observations were conducted once per quarter, for a total of four observations per year. Each observation lasted two to three days and included time spent watching each instructor in the classroom setting and each administrator as she/he provided services to students. Observation activities focused on accumulating evidence related to whether or not, as a school, the model was being implemented as designed. An aggregate summary of school performance was provided at the end of each observation period and reported annually. Feedback on individual teachers was not included in reporting, except to note teachers who were exceptionally effective at implementing each of the core processes described in the model. An example of the reporting for process-context objectives can be found in Table 1.

Table 1
Example of Process-Context Objectives Reporting

PROCESS OBJECTIVE SUMMARY

Model Objectives	On Target for Year 1?	Suggest Improvements for Year 2?
1. Role Models	Yes	Yes
2. Predictable Structure	Yes	No. Currently Meets Year 2 Criteria.
3. Meaningful Relationships	Yes	Yes
4. Challenging Educational Experiences	Suggest Modest Improvement.	Yes
5. Cognitive Dissonance	Yes	Yes
6. Guidance	Yes	Yes

Challenging Educational Experiences Detail: Suggest Modest Improvement

Criteria. Students describe challenging and emotionally intense educational and developmental learning experiences that made a positive impact on them. These activities are usually characterized as doing something they were proud of accomplishing, completing something they previously believed to be beyond their abilities, or demonstrating enthusiasm for trying something they had never done before. Observations include students participating in or preparing for these types of experiences.

Year 1 Findings. Students reported participating in weekly off-campus educational field trips. Further, they described these field trips as helpful, valuable, and something they looked forward to participating in. Additionally, students reported participating in at least one overnight extended learning trip each nine weeks. These trips were described as being challenging, but enjoyable, and students reported feeling proud of what they were able to accomplish while participating in these activities.

Year 2 Suggestions. In addition to remediation and credit recovery, teachers should work together to create collaborative lesson plans and projects designed to build academic self-efficacy and success experiences in school. Examples of these projects include the boat building and regatta project, the Shakespeare scene re-write, and the Florida Keys science exploration trip.

Example 2: Evaluating SLA's short-term student impact objectives. Impact evaluation answers the question "Did we achieve the short-term impacts required to successfully mediate long-term outcomes?" (Chen, 2005; McKenzie et al., 2005). Or, in SLA's case, "Did our students a) adopt the pro-social values, b) set the positive academic and behavioral goals, and c) learn to make the well-reasoned decisions necessary to, d) increase academic and behavioral self-efficacy and promote their success in school?"

To evaluate these objectives, a battery of surveys was administered to all students when they enrolled in SLA, and all SLA graduates were administered the same instrument during their last week at the school. The survey instrument consisted of 13 subscales designed to measure changes in each of the four short-term impact areas: pro-social values, goals, decision making, and self-efficacy. Each of the 13 scales had been widely used in adolescent and school research and reported accepted levels of validity and reliability. Still, prior to use in this evaluation process, the total instrument battery was pilot tested with a group of recent graduates to confirm acceptable levels of validity and reliability within the specific population. A two-week test-retest protocol was used to establish reliability, readability testing was used to determine appropriate instrument reading levels, and each question was presented to and discussed with students to establish thought process validity.

Standard research procedures were used when surveys were administered to ensure high-quality data collection. For instance, all pre-test surveys were administered prior to beginning the treatment (enrollment at SLA); surveys were administered confidentially; and surveys were administered in a quiet, distraction-free environment. Students could voluntarily stop completing the survey at any time, and scores were not used to determine program exit.

At the end of each school year, pre- and post-test scores were compared for all graduating students using dependent sample t-tests. Each sub-scale was tested and reported individually and as related to its corresponding objective. The Type I error rate was set at .10 for all statistical tests. An example of the reporting for short-term student impact results can be found in Table 2.

Example 3: Evaluating SLA's long-term student outcome objectives. Outcome evaluation answers the question "Did we achieve our intended long-term student outcomes?" (Chen, 2005; McKenzie et al., 2005). Or, in SLA's case, "After returning to the traditional school environment, did our students demonstrate intended improvements in academic performance, problem behavior, and attendance rates?"

To evaluate these objectives, SLA used the district's student information system to access and compare:

1. Rates of referrals, suspensions, and expulsions during the nine weeks prior to student enrollment at SLA and during the nine weeks after their placement back in the traditional school setting.
2. Rates of school attendance during the 9 weeks prior to student enrollment at SLA and during the nine weeks after their placement back in the traditional school setting.
3. Student grade point averages during the nine weeks prior to enrollment at SLA and during the nine weeks after their placement back in the traditional school setting.

These results provide important evidence regarding the success of the school as an intervention or treatment. School personnel know students achieve these outcomes while attending SLA—their newly revised student exit system requires students to achieve acceptable levels of each outcome prior to reentering the traditional public school system. However, because successful reentry back into the public school environment stands so central as an SLA's main goal, it was important to track student outcomes in that environment as well. SLA achieves its final success after a student's transition back into the public school environment; therefore, monitoring post-transition outcomes is critically important.

Additionally, school personnel administered reading and math evaluations during the first and last weeks of enrollment at SLA to determine corresponding increases or decreases in student reading and math levels while attending the school. An example of the reporting for long-term student outcomes can be found in Table 3.

Other evaluation considerations. Because schools are composed of a set of complex and dynamic human systems, evaluations that attempt to reduce a school to a set of quantitative scores are bound to provide an incomplete and, possibly, insulting picture of school performance. Most school evaluations require the use of both quantitative and qualitative methods. At a minimum, most school evaluation reports require qualitative descriptions of the school, the district, and the community context in which the evaluation is being conducted.

Table 2
Examples of Short-term Student Impact Reporting

VALUES DETAIL

Instrument	Pre-test Mean	Post-test Mean	Significant Difference?
Marlowe-Crowne Social Desirability Scale	83.89	91.84	Yes (.07)
PALS Academic Values-Content	23.56	23.13	No (.66)
PALS Cheating, Disruptive Behavior & Skepticism about the Relevance of School	28.62	33.93	Yes (.03)
		Total Impact	**Yes**

Pre-post test data suggest the SLA program made a real and positive difference in program graduates' commitment to pro-social values. Students were more likely to agree with a range of pro-social items including "I am always willing to admit it when I make a mistake," "I don't find it particularly difficult to get along with loud-mouthed, obnoxious people," "I don't mind admitting that I don't know something," and "I am always courteous and polite even to people who disagree with me." Even though sample sizes were relatively small, the differences in the pre-post surveys were large enough to achieve statistical significance and provide reasonable evidence that the program is contributing to student improvements in this area.

Pre-post test data also suggest the SLA program might be less effective at reinforcing values related to academic content. Survey items included "I believe it is important to have good reading skills," "I believe studying reading is interesting," "I believe it is important to have good math skills," and "I believe studying math is interesting." Small sample sizes, however, make it difficult to draw a definitive conclusion. For now, faculty and staff should be aware of this as a possible area for improvement.

Pre-post data also suggest the SLA program was effective at helping students adopt values that preclude cheating and disruptive behavior and include belief that school is relevant and helpful. Survey items included "Even if I do well in school, it will not help me have the kind of life I want when I grow up," "I sometimes annoy my teacher during class," "Doing well in school doesn't improve my chances of having a good life when I grow-up," and "I sometimes copy answers from other students during tests." Graduating students were less like to agree with these types of statements.

GOALS DETAIL

Instrument	Pre-test Mean	Post-test Mean	Significant Difference?
The Hope Scale	40.87	43.33	Yes (.00)
Importance Assessment of Academic Goals	30.98	33.82	Yes (.09)
PALS Mastery & Performance vs. Avoidance Goal Orientations Subscale	30.80	32.67	Yes (.02)
		Total Impact	**Yes**

Pre-post data suggest SLA makes its most significant difference in the area of goal-orientation. In all 3 subscales, evidence suggests a significant difference in overall goal orientation, academic goal orientation, and willingness to apply effort to identified goals. Example survey items include "I energetically pursue my goals" "I meet the goals that I set for myself," "Even when others get discouraged, I know I can find a way to solve the problem," "It is important to me to pay attention during class," "It is important to me to work hard in all my classes," "One of my goals in class is to learn as much as I can," and "One of my goals is to master a lot of new skills this year."

The core of the SLA system focuses on helping students work toward relevant and authentic goals. These goals emphasize program completion, school success, and positive life outcomes. Evidence suggests the process of pursuing these goals in the context of the SLA environment is producing positive results in this area.

Table 3
Example of Long-term Student Outcome Reporting

PROBLEM BEHAVIOR DETAIL: 9 WEEKS PRIOR TO SLA COMPARED TO 9 WEEKS AFTER SLA

Criteria	Pre-program Mean	Post-program Mean	Significant Difference?
Referrals	6.77	2.38	Yes (.00)
Suspensions (in days)	1.81	0.25	Yes (.02)
Expulsions	0.69	0.00	Yes (.06)
		Total Impact	

Attendance Detail: 9 Weeks Prior to SLA Compared to 9 Weeks After SLA

Criteria	Pre-program Mean	Post-program Mean	Significant Difference?
Excused Absences	3.33	1.43	Yes (.01)
Unexcused Absences	7.20	1.72	Yes (.00)
All Absences	10.51	3.18	Yes (.00)
		Total Impact	

Academic Improvement Detail: 9 Weeks Prior to SLA Compared to 9 Weeks After SLA

Criteria	Pre-program Mean	Post-program Mean	Significant Difference?
Grade Point Average	1.40	2.84	Yes (.04)
		Total Impact	

Academic Improvement Detail: First Week of SLA Compared to Last Week at SLA

Criteria	Pre-program Mean	Post-program Mean	Significant Difference?
Reading Achievement	4.47	6.17	Yes (.05)
Math Achievement	4.92	7.08	Yes (.03)
		Total Impact	

Phase 4: Using evaluation data to promote rigorous practice

Each year SLA compiles all of their evaluation results into an annual report. A draft is completed no longer than 14 days after the last day of school, and a copy is e-mailed to all school personnel. All team members have seven days to make comments or suggest revisions to the report. Report revisions are finalized within the next seven days, and the final report is e-mailed to all school personnel. All team members receive a copy of the final annual evaluation report within 30 days of the last day of school. The final report includes a description of the school, the district, and the community it serves; a brief description of the *Model of Student Empowerment for Alternative Middle Schools*; a list of specific strategies used to support the model's process-context objectives; a description of all evaluation methods; and all evaluation findings for the year, including recommendations for the coming year.

When school resumes, back-to-school planning begins with a review and discussion of these findings. Both the model and the previous year's evaluation results are central to shaping and guiding plans for the coming year. All annual planning and goal documents are formatted to match the components of

the model. Further, each year, specific strategies are developed for each model process, and goals are set for each student outcome. Training for new personnel begins with the model and includes all previous years' evaluation data.

Each year, school administrators and the author reevaluate the model itself. All data is examined to determine whether the model accurately represented the relationships between each of the processes, impacts, and outcomes presented. At the end of Year 3, extensive statistical modeling will be conducted to determine the relative influence of each model component and its predictive value. After this analysis, decisions will be made regarding how to best refine the model. If necessary, any additions, subtractions, or revisions to the current model will be made at that time.

Finally, educators dedicated to the highest levels of rigor are required to extend themselves beyond boundaries of the local school. Any model can benefit from other educators' expertise, experience, and professional scrutiny. As such, presenting and publishing both the model and the ensuing evaluation findings represents an important step in the SLA evaluation plan.

Discussion

The education profession stands at a crossroads of critical policy decisions related to school evaluation. A range of arguments can be made about the relative advantages and disadvantages of NCLB. Educators can cite valid examples of NCLB both helping and harming (Griffin & Scharmann, 2008; Schmidt, 2008; Stanik, 2007). Ultimately, the question becomes, "How do we know when these trade-offs are acceptable and when they are not?" In this particular case, SLA administrators decided they could not afford to divert too many of their resources directly to NCLB objectives. Consistent with the middle school model, administrators believed preserving certain school contexts and processes devoted to the development of the whole child were essential to improving student reading and math scores. Further, they believed that whole-child development made critically important contributions to their students' lifelong ability to do something positive and productive with their reading and math skills. The significance of this contribution should not be overlooked. Although the quantity and quality of direct instructional time designated to reading and math can make a critical contribution to positive student outcomes, so can the broader goals of helping students be ready to learn; motivating them to achieve; and meeting their basic needs in a manner that supports intellectual curiosity and the hunger to pursue education, career, and life goals.

In no way does this case study suggest national evaluation efforts be expanded to include any of these additional priorities or services. It does suggest that sometimes schools and districts benefit from having a counterweight to NCLB—a way of making sure other priorities receive reasonable attention, adequate funding, and administrative support. Further, it suggests that properly supported and motivated educators can be trusted to identify and pursue these types of priorities at the local level. In a diverse nation, one approach will never adequately fit the needs of every child, school, or community. In these cases, local schools and districts can use similar evaluation models to successfully pursue objectives outside the scope of national or state initiatives.

This case study also highlights the influence of systems, both locally and nationally. Recently, much of the national discussion about education has focused on holding underperforming teachers accountable. Ineffective teachers unwilling or unable to perform effectively with students should not be allowed to continue in the classroom; however, neither should ineffective or underperforming systems survive without scrutiny, revision, and harmony with best practices. System-level problems can and do undermine school and teacher effectiveness and student performance (Ervin, Schaughency, Matthews, Goodman, & McGlinchey, 2007; McIntosh, Filter, Bennett, Ryan, & Sugai, 2010). This case study supports the efficacy of using evaluation strategies to identify system-level problems in local schools and to realign system-level efforts to better support successful educator and student outcomes.

Finally, although the example provided focuses on an alternative middle school, the processes described relating to modeling, systems review, and evaluation can be applied to any type of middle school. Every middle school has distinctive ways of working with students that are worth preserving and unique goals worth achieving.

Conclusions

Middle school educators want their students to do well. They want their schools and classrooms to be places where young people learn and grow. They want to take pride in their work, knowing their efforts make a real difference. In its truest form, evaluation

provides educators with the assurance that these things are happening consistently, for all students, and at a high level (Chen, 2005; McKenzie et al., 2005). If, for some reason, they are not, evaluation reveals this fact and compels action. As a result, evaluation can be an important tool that, when used properly, helps an educator proceed with the confidence to face any situation (Chen, 2005; McKenzie et al., 2005).

Sadly, evaluation sometimes feels less like an educator's tool and more like a taskmaster (Schoen & Fusarelli, 2008). This tendency does not mean that evaluation efforts should be avoided, abandoned, or minimized. If anything, it demands that educators take more leadership in this area of practice—that they actively reclaim the benefits of school evaluation.

One way to demonstrate this leadership is by developing positive and effective evaluation models. This case study describes the attempt of one middle school to do just that—to fully embrace evaluation and to exemplify how to use it to effectively promote rigor-in-practice. In this case, the featured middle school used evaluation to clarify their goals and how they do their work; to align all of the school's efforts to achieve those goals; and to measure, track, and plan for continual progress. They chose to do this in a collaborative, transparent, and positive way that emphasized promoting growth by thinking deeply and systematically about the needs of their students and the work of their school, and by using rigorous evaluation methods to help them identify, celebrate, and disseminate their successes. This case study offers an approach that replaces externally-motivated evaluation cultures—where educators work only to avoid the threat of punishment—with an internally-motivated evaluation culture, in which educators grow increasingly committed to "always getting better, because we are professionals that care about our students and our work."

References

Abreu, A. A., Macedo, P. P., & Camarinha-Matos, L. M. (2009). Elements of a methodology to assess the alignment of core-values in collaborative networks. *International Journal of Production Research, 47*(17), 4907–4934.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 94–107.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 212–234.

Barnstable, R., Cargill, L., Gehlbach, S., & Workman, H. (1997). Advancing the Pro-Social Skills in At-Risk Elementary Students through Curricular Interventions. Student research, Saint Xavier University, Chicago, IL. Retrieved from EBSCOhost.

Chen, H. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness.* Thousand Oaks, CA: Thousand Oaks Press.

Dahl, R. E. (2001). Affect regulation, brain development, and behavioral/emotional health in adolescence. *CNS Spectrum, 6*, 1–12.

Ervin, R. A., Schaughency, E., Matthews, A., Goodman, S. D., & McGlinchey, M. T. (2007). Primary and secondary prevention of behavior difficulties: Developing a data-informed problem-solving model to guide decision making at a school-wide level. *Psychology in the Schools, 44*(1), 7–18.

Forgas, J. P. (2001). *Handbook of affect and social cognition.* New York, NY: Lawrence Erlbaum.

Glasser, W. (1999). *Choice theory: A new psychology of personal freedom.* New York, NY: HarperCollins.

Keating, D. P. (2004). Cognitive and brain development. In Lerner, R. J. & Steinberg, L. D. (Eds.) *Handbook of adolescent psychology* (pp. 45–84). Chicago, IL: Wiley.

Kennedy, D. M., & McComb, S. A. (2010). Merging internal and external processes: Examining the mental model convergence process through team communication. *Theoretical Issues in Ergonomics Science, 11*(4), 340–358.

Lewis, A. (2007). Open to the public: How communities, parents, and students assess the impact of the "No Child Left Behind Act," 2004–2007. *Public Education Network.*

Mann, M. (2010). Sunshine Learning Academy evaluation report for July 3rd, 2010. Unpublished report.

McClure, L., Yonezawa, S., & Jones, M. (2010). Can school structures improve teacher-student relationships: The relationship between advisory programs, personalization and students' academic achievement. *Education Policy Analysis Archives, 18*(17), 47–61.

McIntosh, K., Filter, K. J., Bennett, J. L., Ryan, C., & Sugai, G. (2010). Principles of sustainable prevention: Designing scale-up of school-wide positive behavior support to promote durable systems. *Psychology in the Schools, 47*(1), 5–21.

McKenzie, J. F., Neigler, B. L., & Smeltzer, J. L. (2005). Planning, implementing, and evaluating health promotions programs. San Francisco, CA: Benjamin Cummings.

Miller-Williams, S. L., & Kritsonis, W. (2009). A systems approach to comprehensive school reform: Using the realms of meaning and the Baldridge Approach as a systems framework. Online submission. *National Forum of Applied Educational Research Journal, 23*(1–2). Retrieved from EBSCOhost.

Moriarty, B., Douglas, G., Punch, K., & Hattie, J. (1995). The importance of self-efficacy as a mediating variable between learning environments and achievement. *British Journal of Educational Psychology, 65*, 73–84.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*, 30–38.

Nowak, A. (2004). Dynamical minimalism: Why less is more in psychology. P*ersonality & Social Psychology Review, 8*(2), 183–192.

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology, 24,* 124–139.

Resick, C. J., Murase, T., Bedwell, W. L., Sanz, E., Jiménez, M., & DeChurch, L. A. (2010). Mental model metrics and team adaptability: A multi-facet multi-method examination. *Group Dynamics: Theory, Research, and Practice, 14*(4), 332–349.

Schmidt, T. (2008). Scratching the surface of "No Child Left Behind": How "No Child Left Behind" unfairly affects schools with significant proportions of disadvantaged students. Online submission. Student research project, Dominican University of California. Retrieved from EBSCO*host.*

Schoen, L., & Fusarelli, L. D. (2008). Innovation, NCLB, and the fear factor: The challenge of leading 21st-Century Schools in an era of accountability*. Educational Policy, 22*(1), 181–203.

Sowell, E.R. et al. (2002). Development of cortical and subcortical brain structures in childhood and adolescence: A structural MRI study. *Developmental Medicine and Child Neurology, 44*, 4–16.

Taylor, B. (2009). Classroom management impacts student achievement: Tips to thrive and survive. Online submission. Student research, Jackson State University, Jackson, MS. Retrieved from EBSCO*host.*

